JOC The Journal of Organic Chemistry

Article

pubs.acs.org/joc

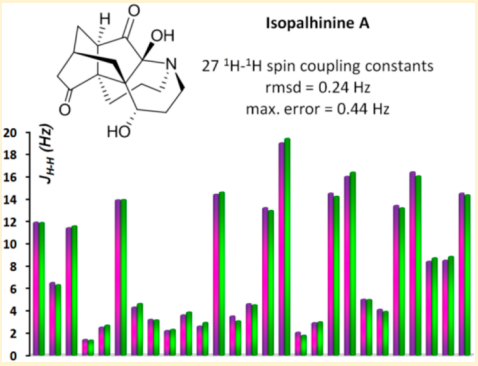
Relativistic Force Field: Parametric Computations of Proton−Proton Coupling Constants in ^{1}H NMR Spectra

Andrei G. Kutateladze* and Olga A. Mukhina

Department of Chemistry and Biochemistry, University of Denver, Denver, Colorado 80208, United States

Supporting Information

ABSTRACT: Spin−spin coupling constants in ^{1}H NMR carry a wealth of structural information and offer a powerful tool for deciphering molecular structures. However, accurate ab initio or DFT calculations of spin−spin coupling constants have been very challenging and expensive. Scaling of (easy) Fermi contacts, fc, especially in the context of recent findings by Bally and Rablen (Bally, T.; Rablen, P. R. *J. Org. Chem.* **2011**, *76*, 4818), offers a framework for achieving practical evaluation of spin−spin coupling constants. We report a faster and more precise parametrization approach utilizing a new basis set for hydrogen atoms optimized in conjunction with (i) inexpensive B3LYP/6-31G(d) molecular geometries, (ii) inexpensive 4-31G basis set for carbon atoms in fc calculations, and (iii) individual parametrization for different atom types/hybridizations, not unlike a force field in molecular mechanics, but designed for the fc's. With the training set of 608 experimental constants we achieved rmsd <0.19 Hz. The methodology performs very well as we illustrate with a set of complex organic natural products, including strychnine (rmsd 0.19 Hz), morphine (rmsd 0.24 Hz), etc. This precision is achieved with much shorter computational times: accurate spin−spin coupling constants for the two conformers of strychnine were computed in parallel on two 16-core nodes of a Linux cluster within 10 min.



1. INTRODUCTION

NMR is unquestionably the most informative solution structure method. Proton spin−spin coupling constants (SSCC) in ^{1}H NMR carry a wealth of structural information and offer a powerful and intuitive predictive tool. Yet, until recently, the majority of computational papers on NMR predictions relied on matching the experimental and computed NMR chemical shifts, not the spin coupling constant, to rule out incorrect structures and endorse the ones fitting best. This is not ideal because predicted chemical shifts carry virtually no structural information content and therefore offer little help in informing and guiding the process of structural assignment.[1] Alas, accurate calculations of spin−spin coupling constants still present a formidable challenge and are computationally expensive, as they require high-quality correlated wave functions, large basis sets, and evaluation of multiple relativistic mechanisms mediating the interactions between nuclear spins, most importantly: (i) dia- and paramagnetic components of spin−orbit coupling and (ii) hyperfine coupling with contributions from local, i.e., Fermi contacts (fc), and nonlocal spin-dipole interactions.[2] The problem is aggravated by the fact that in conformationally flexible systems one needs to average the results over an ensemble of conformers computed separately. This further accentuates the need for a fast and reliable method of SSCC calculations, capable of handling multiple conformers expeditiously.

In this context, an alternative approach for computing the spin−spin coupling constants via linear scaling of Fermi contacts is gaining traction. During the last two decades, there has been a growing consensus that the Fermi contact mechanism dominates nuclear spin scalar couplings.[3] There is also a general understanding that a basis set with additional tight s-type functions is critical for calculating accurate fc's.[4]

Bally and Rablen[5] capitalized on all these relatively recent advances and proposed a very practical methodology for calculations and linear corrections of Fermi contacts to estimate SSCCs, which involved an uncontracted basis set on hydrogen atoms augmented with such a tight 1s-type function, which they term (u + 1s). We refer the reader to their excellent paper for an overview of the state of the field and analysis of the difficulties involved. It is also worth drawing the reader's attention to one of their more general statements: "Insofar as the practice of predicting ^{1}H NMR spectra using electronic structure theory becomes commonplace even among non-specialists, as we believe it should, it would certainly behoove practitioners to know what combinations of methods, functionals, basis sets, etc., provide the best balance between cost and accuracy." Needless to say, we fully share this philosophy.

Here, we report that utilization of an uncontracted basis set augmented with an additional 1s function for fc calculations can be systematically improved contingent on two assumptions: (i) different hybridization types require different linear scaling parameters and (ii) such parametrization requires non-zero

Received: August 2, 2014
Published: August 26, 2014

   dx.doi.org/10.1021/jo501781b | J. Org. Chem. 2014, 79, 8397−8406

intercepts but can be made stable and accurate provided that a large number of experimental constants is used for the training set in each linear correlation.

The use of three different scaling factors for vicinal constants of the three different hybridization combinations, i.e., sp^3–sp^3, sp^3–sp^2, and sp–sp^2, was not unprecedented, although it is not entirely clear to us why Pereda-Miranda et al.[6] elected to parametrize the results of high-end computations involving all the difficult contributions to SSCC. It is surely more practical to parametrize the truncated low-level computations to achieve precision at low cost.

A rigorous paper by Weinhold, Markley, et al.[7] on natural *J*-coupling analysis, i.e., the interpretation of scalar *J*-couplings in terms of natural bond orbitals (NBOs), helped clarify the nuances of applying Weinhold's NBO analysis[8] to evaluation of Fermi contacts. A similar analysis involving slightly different semilocalized MOs was proposed by Perlata, Contreras, and Snyder.[9]

In our opinion, all these findings encourage and validate the timely development of the methodology put forward in our study.

2. RESULTS AND DISCUSSION

Using the NBO program[10] as implemented in the Gaussian 09 computational package, we have calculated the results of NBO analysis for Bally and Rablen's training and probe sets of compounds and attempted to correlate the residuals, i.e., (J_{exp} − fc), with several NBO parameters related to carbons' hybrid orbitals: the pairwise product of C–(H) hybridizations, pre-NHO overlaps, and the energies derived from the second-order perturbation analysis of the interacting donor and acceptor NBOs. The last term was central to Weinhold's discussion of "how *J*-coupling, or rather the transfer of spin density, is related to spin hyperconjugative delocalization (by means of second order perturbation analysis)...".[7] We found that most of these NBO values showed modest correlation with the residuals (J_{exp} − fc), with correlation coefficients ranging from 0.41 to 0.49, except for the pre-NHO overlap, which showed negative correlation.

From this result we drew two conclusions: (1) the calculated fc's could potentially be corrected for different hybridization states of the connected carbons, and (2) while quantitative corrections using computed NBO parameters somewhat improve the accuracy of SSCCs calculations, it was abundantly clear that such improvement was very modest. It was concluded at this point that a more practical approach to fast and accurate calculations of SSCCs should involve: (i) selection and individual parametrization of a representative set of SSCC "types" defined by connectivity and hybridization (for example, geminal sp^2, or vicinal sp^2–sp^3, etc.), i.e., not unlike the parametrization (force fields) in molecular mechanics, and (ii) optimization of an uncontracted basis set for hydrogen, augmented with a tight 1s-type function, in an iterative procedure in which the individual scaling factors for respective types are optimized simultaneously with the basis set optimization.

2.1. Parametrization and Basis Set. Successful implementation of this strategy is predicated upon utilization of a much larger training set allowing for several data points per type. It is important because we allow for non-zero intercepts in our linear scaling which, at a first glance, may conflict with the assertion by Bally and Rablen that "there is no theoretical justification for a nonzero intercept".[5] While this may be true for an infinite set of all possible variations of structure and hybridization, the omission of spin–orbit and spin–dipole components *for a given structural and hybridization subtype* introduces a systematic error, which can be corrected by a non-zero intercept. To illustrate this point, Figure 1 shows a linear

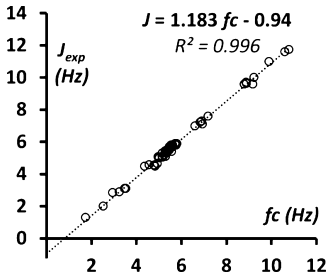


Figure 1. Linear correlation obtained for a set of 52 sp^2–sp^2 (*cis*-ene) SSCCs, rmsd = 0.14 Hz.

correlation for the *cis*-ene type SSCC, where the rmsd (root-mean-square deviation) of 0.14 Hz was achieved using a set of 52 experimental constants with the absolute intercept value of 0.94 Hz exceeding the rms deviation for the set by almost 7-fold. This leaves little doubt that the intercepts are necessary for accurate subtype parametrizations.

As we have shown in the past, the spin–orbit coupling itself (in organic triplet diradicals) can be parametrized,[11] as indeed could be the case for other *difficult* relativistic components in the SSCC computations. However, an excellent correlation between the calculated Fermi contacts and experimental spin–spin coupling constants shown in Figure 1 implied that a well-parametrized set of structural and hybridization types may suffice in achieving very high accuracy of SSCC calculations based solely on the computed fc's.

Our full training set of 608 proton spin–spin coupling constants was assembled using data available in the literature and the spectral database of organic compounds by AIST, Japan.[12] In order to meet another goal for this methodology, i.e., to achieve considerable reduction of computational time, we implemented the following: (i) a minimalist level of DFT theory, B3LYP/6-31G(d), was employed for structure optimizations, and (ii) the basis set for hydrogens was optimized in conjunction with a very inexpensive basis set, 4-31G, for carbon atoms, 6-31G for Li, Be, B, N, O, and F, and finally, 3-21G* for Si, S, P, As, Se, Br, Cl, Br, and I. An iterative procedure for the basis set optimization was continued until the rmsd for the full set of 608 experimental constants reached 0.186 Hz and was not decreasing any further. Table 1 compares the exponents for the obtained basis set (which we term DU4) with that of Bally and Rablen.

With two exceptions, our optimization generally tightened the s-type exponents. The exponent for the p-type function was somewhat decreased.

The optimization of the basis set was carried out simultaneously with refinement of the linear scaling. Figure 2 shows the SSCC types and their respective scaling parameters, which gave the best overall rmsd value in conjunction with the optimized basis set.

The number of experimental data points per type varied from more than 70 in the case of vicinal saturated fragments to low 10s. However, three of these parameters were based on less than 10 experimental data points each. For two of them, as a precaution, we chose to set the intercept to zero until more

dx.doi.org/10.1021/jo501781b | J. Org. Chem. 2014, 79, 8397–8406

Table 1. Basis Sets Compared

type	DU4 (this work)	ref 5	% ratio
S	14.573150	18.731137	77.8
S	3.968624	2.825394	140.4
S	0.352791	0.640122	55.1
S	0.180730	0.161278	112.0
S	240.390929	56.193411	427.7
S	385.735553	168.580233	228.8
S	1685.684542	505.740698	333.3
S	2387.167390	1517.222094	157.3
P	0.491090	1.100000	44.6

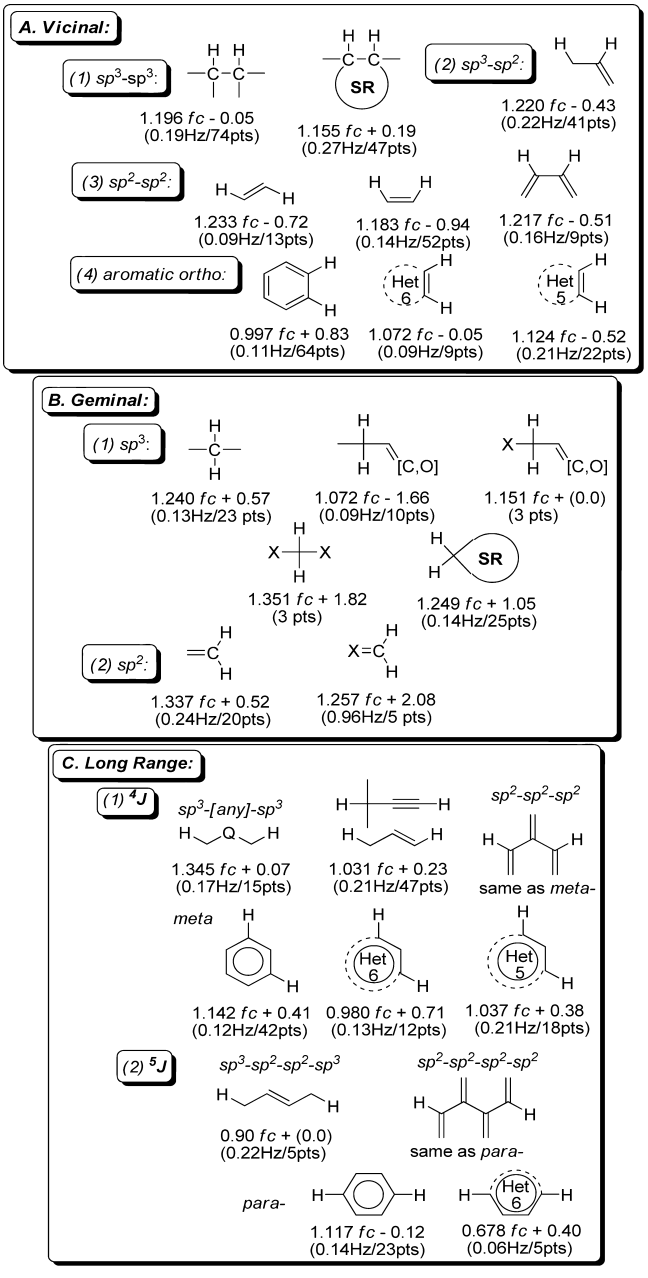


Figure 2. SSCC types and the linear scaling parameters developed for the fc's. The rmsd values and the number of experimental points for individual types are shown in parentheses. SR = small (3,4) rings.

experimental constants become available to confidently determine the values for both the slope and the intercept.

The achieved considerable improvement of the accuracy as measured by rmsd <0.19 Hz is not the only advantage of our methodology. The (related) small number of outliers is even more critical for confident structural assignment. As it is evident from Figure 3 the absolute error distribution is very narrow.

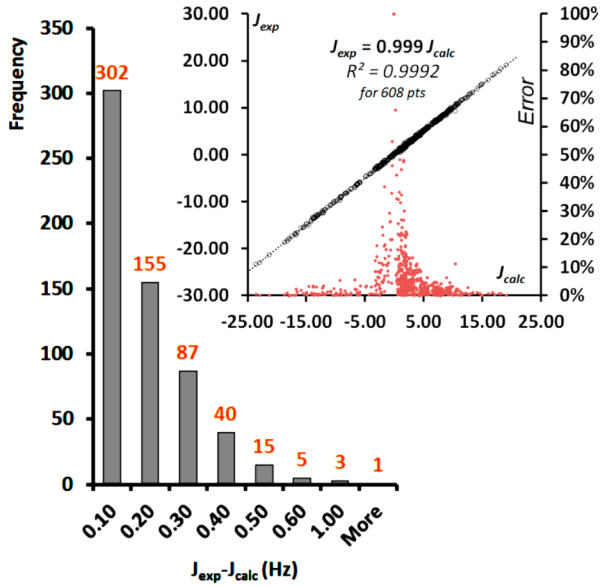


Figure 3. Unsigned error distribution histogram shows only 9 out of 608 data points deviating by more than 0.5 Hz. The inset shows a near-perfect linear correlation plot of computed and experimental constants (black empty circles) and small relative errors (red circles, secondary axis, %), which become significant only for very small SSCCs.

Only nine points out of 608 deviated by more than 0.5 Hz, with only one of them deviating by more than 1 Hz (1.05 Hz, sp^3–sp^3 *cis*-vicinal constant in strained cyclobutene).

This *tightening* of the error distribution curve imparts certainty that the SSCCs computed with our "relativistic force field" are accurate enough to unambiguously predict structures or confidently rule out the misassigned ones.

The technical implementation details are described in the Supporting Information. We took advantage of OpenBabel's[13] Perl bindings and created scripts to auto assign the SSCC types shown in Figure 2 while scaling the computed Fermi contacts accordingly. Using examples of complex organic polycyclic compounds, we now will show that the method performs very well in terms of both accuracy and computation time.

2.2. Conformationally Rigid Test Cases. Bis-oxetane **A**, Figure 4, described in our previous work,[14] offered a complex rigid test model that did not require conformational averaging. Calculated SSCCs for this C_i-symmetric structure matched the experimental constants exceptionally well, rmsd = 0.16 Hz and maximum unsigned error (mue) 0.26 Hz over the full set of 17 symmetry unique constants. Figure 5 illustrates how well its multiplets are simulated with the computed SSCCs.

For comparison, we calculated SSCCs for bis-oxetane A using Bally and Rablen's approach[5] and obtained rmsd of 0.51 Hz and a maximum unsigned error of 1.21 Hz, Table 2. The rmsd is more than three times that of the value obtained with our relativistic force field parametrization. However, the rather large maximum deviation of 1.21 Hz points to a bigger potential problem. Significant overlap of computed SSCC values for hypothetical structures under consideration may

dx.doi.org/10.1021/jo501781b | J. Org. Chem. 2014, 79, 8397–8406

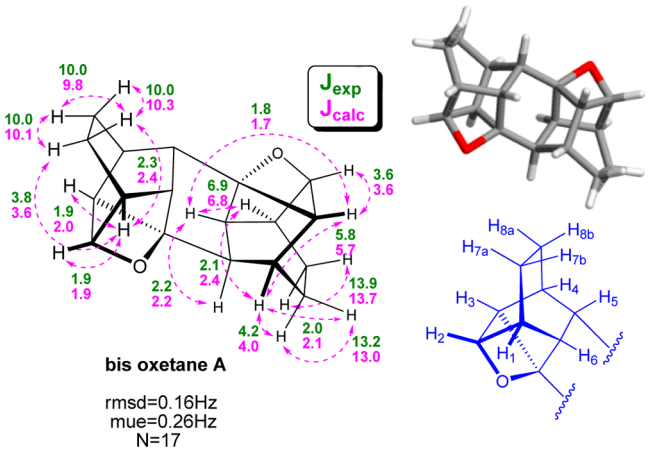


Figure 4. Rigid bis-oxetane A. [Left] experimental (green) and calculated (maroon) SSCCs mapped on its drawing. [Right] its B3LYP/6-31G(d) structure (top); proton numbering for Table 2 (bottom). Here and later in the text the absolute SSCC values are shown.

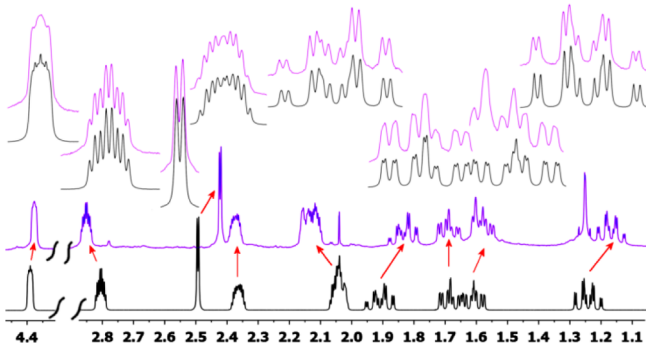


Figure 5. Experimental (top) and simulated NMR spectra for bis-oxetane A; linearly corrected chemical shifts were separately computed with the mPW1PW91 functional.[15]

Table 2. Bis-oxetane A: Comparison of Calculated SSCCs[a]

		method			
		this work		Rablen–Bally, ref 5	
H_x–H_y	J_{exp}	J_{calc}	ΔJ	J_{calc}	ΔJ
1–2	5.8	5.65	0.15	5.65	0.15
1–5	2.1	2.36	−0.26	2.51	−0.41
2–3	3.6	3.63	−0.03	3.73	−0.13
2–4	1.9	1.95	−0.05	1.74	0.16
2–5	1.8	1.7	0.10	1.43	0.37
3–4	1.9	1.87	0.03	1.94	−0.04
4–5	6.9	6.81	0.09	7.66	**−0.76**
5–6	2.2	2.16	0.04	2.40	−0.20
1–7a	2	2.12	−0.12	2.06	−0.06
1–7b	4.2	3.99	0.21	4.38	−0.18
4–8a	3.8	3.56	0.24	3.57	0.23
4–8b	2.3	2.39	−0.09	2.48	−0.18
7a–7b	13.2	12.98	0.22	12.20	**1.00**
7a–8a	10	10.14	−0.14	10.46	−0.46
7a–8b	10	9.79	0.21	10.51	**−0.51**
7b–8b	10	10.26	−0.26	10.67	**−0.67**
8a–8b	13.9	13.71	0.19	12.69	**1.21**

[a]Absolute deviations greater than 0.5 Hz are shown in bold type.

make or break confident assignment (or ruling out) of a candidate structure. *Timewise, these less accurate results took two times longer to compute*: 14 min vs 7 min wall time on a 16-core node of a Beowulf cluster. The acceleration offered by our *more accurate* method is significant, especially because one often deals with a large ensemble of contributing conformers.

Two additional examples of rigid polycyclic compounds, recently reported by Porco, Stephenson, et al.,[16] are shown in Figure 6. Five reported spin–spin coupling constants for

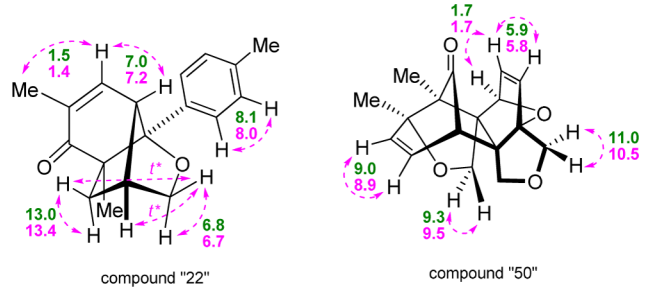


Figure 6. Structures from ref 16. For the apparent triplet designated with *t** in compound **22**, see the text.

compound **50** were calculated with rmsd of 0.26 Hz. For the seven experimental constants reported for compound **22**, we obtained rmsd = 0.28 Hz. However, the largest deviations were associated with the apparent triplet, J = 2.6 Hz, denoted by *t** in Figure 6. This is a typical practice in organic literature where doublets of doublets, with constants differing by 1 Hz or less, are generally reported as apparent triplets. In this particular case, if we assumed that 5.2 Hz, i.e., double the experimental value, accurately represents the sum of the two actual constants, then rmsd improves to 0.19 Hz.

Corey's 1988 paper on the synthesis of a complex hexacyclic trilactone (±)-ginkgolide B[17] listed six experimental proton constants, for which the calculated values gave rmsd of 0.21 Hz with maximum unsigned error of 0.34 Hz, Figure 7. This

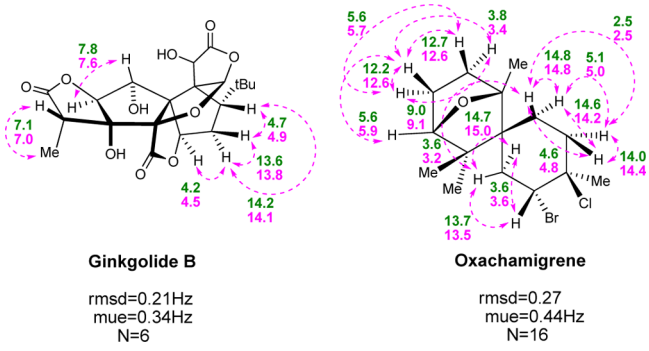


Figure 7. Ginkgolide B and oxachamigrene.

example underscores the superiority of SSCCs calculations over chemical shifts: as the 500 MHz spectrum of ginkgolide B was recorded in aqueous acetone the chemical shifts prediction suffered but not the calculated spin–spin coupling constants. The calculated SSCCs for a given structure are generally less sensitive to solvent effects. More profound solvent effects on the spin-coupling constants are often exerted in conformationally flexible systems through the change in conformers' content at equilibrium.

dx.doi.org/10.1021/jo501781b | J. Org. Chem. 2014, 79, 8397–8406

Oxachamigrene,[18] a halogenated sesquiterpene from *Laurencia obtusa*, gave us 16 accurate experimental constants to work with. The rmsd of 0.27 Hz and mue of 0.44 Hz did not leave any doubt that its structure was assigned correctly (Figure 7, right).

The structure of the photodimer of another sesquiterpene, chloranthalactone A, was initially assigned incorrectly but was subsequently corrected on the basis of additional spectroscopic data.[19] SSCCs for chloranthalactone A were predicted with rmsd 0.32 Hz and with maximum deviation of 0.67 Hz, which is plausibly caused by the reported apparent triplet of 7.3 Hz as a part of the multiplet for cyclopropyl's C(1)−H (see numbering in the inset of Figure 8). The corrected structure for the photodimer gave rmsd of 0.1 ppm for chemical shifts and 0.43 Hz for spin−spin coupling constants. The same apparent triplet problem was a suspect here. If the apparent triplet for C(1)−H is treated as a sum of two doublets as described above, rmsd improves to 0.33 Hz.

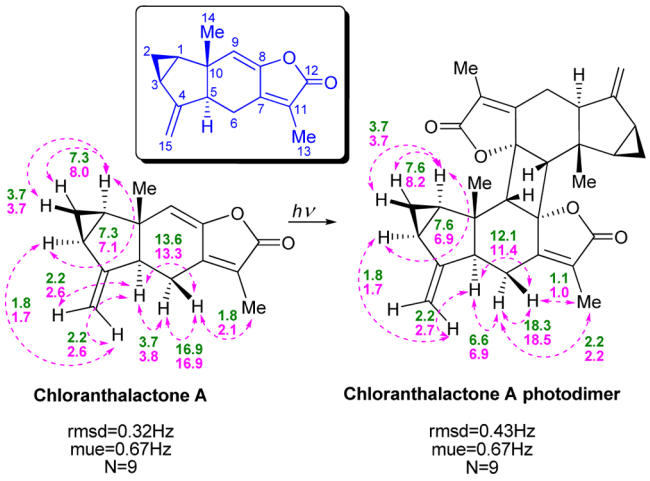


Figure 8. Chloranthalactone A and its photodimer.

We calculated NMR spectra for all four possible [2 + 2] dimers: head-to-head/head-to-tail and syn/anti. Two of the remaining three isomers, anti-HH and syn-HT, showed severely mismatched constants with error exceeding 4 Hz and were discarded. However, the syn-HT isomer showed rmsd of 0.53 Hz for SSCCs and 0.28 ppm for chemical shifts. The small difference between rms deviations of 0.53 and 0.43 Hz gives an instructive example that very accurate computations of proton spin−spin coupling constants can help rule out an incorrect structure even in the situation when the rmsd difference is only 0.1 Hz. The calculated chemical shifts also confirm the selection of the correct structure.

We then validated our methodology with more challenging structures, which generally required some conformational averaging.

2.3. Cases Requiring Conformational Averaging. Strychnine, Figure 9, is unique in the sense that it has been a focus of a number of NMR studies, and its spin−spin coupling constants have been measured with due diligence.[20] Once the structures of the two conformers were optimized at an inexpensive B3LYP/6-31G(d) level of theory (each took approximately 10 min wall time on a 16-core node of a Beowulf cluster), the actual computations of accurate spin−spin coupling constants took 9 min of the wall time and resulted in

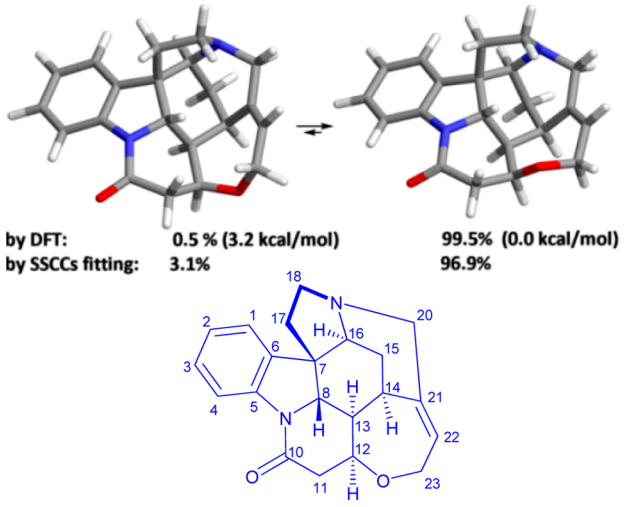


Figure 9. (Top) Structures of two conformers of strychnine and their percent content at equilibrium based on either B3LYP/6-311+G(d,p)//6-31G(d) energies or fitting of the computed SSCCs. (Bottom) Numbering scheme for Table 2

rmsd of 0.19 Hz, with maximum deviation of 0.48 Hz (Table 3).

If the calculated content of strychnine conformers at equilibrium is based on the B3LYP/6-311+G(d,p)//6-31G(d) gas-phase energy difference, Figure 9, the conformer ratio of 99.5% to 0.5% is obtained. A polarized continuum model to account for solvent (IPCM, chloroform) changes this ratio negligibly in a wrong direction, 99.6−0.4%. It is instructive that the value of 3.1% for the minor conformer, obtained by weighing our calculated SSCCs to match the experimental constants, *is in much better agreement with the recently reported 2.5% value experimentally measured by high precision NOE.*[21] This result offers an alternative look at the accurate determination of conformer ratios by fitting the weighted averages of the properly scaled Fermi contacts for individual conformers with the experimental constants. We maintain that the distribution of conformers at equilibrium is better evaluated by such fitting of the calculated spin−spin coupling constants, not the DFT-based Boltzmann populations. This is particularly true for the cases when the constants vary significantly for individual conformers, offering a large dynamic range for accurate fitting.

Computational time is always an issue for a multiconformer system, in which case additional potential "savings" could be considered: the full matrix of Fermi contacts need not be computed. Often, due to overlapping multiplets and other complications, only a small number of experimental SSCCs are measured and reported. Thus, further acceleration can be readily achieved by excluding protons, for which experimental constants are not known, from the fc computations.

We next present several additional examples where accurate prediction of experimental constants, obtained by conformational averaging, was generally commensurate with the relative DFT energies of the contributing conformers. The first example is our previously reported product of intramolecular [4 + 4] cycloaddition of azaxylylene, photogenerated from furylpropanamidoindanone,[22] which gave rmsd = 0.26 Hz and mue = 0.47 Hz, Figure 10a. These values are based on averaging over two conformers defined by conformational mobility in the five-membered lactam ring (for the actual structures of conformers,

dx.doi.org/10.1021/jo501781b | J. Org. Chem. 2014, 79, 8397−8406

Table 3. SSCCs (Hz) for Strychnine Protons[a]

H_x–H_y	J_{exp}	J_{calc}	ΔJ	H_x–H_y	J_{exp}	J_{calc}	ΔJ
1–2	7.35	7.54	−0.19	14–15b	4.77	4.70	0.07
1–3	1.41	1.30	0.11	14–22	2.82	2.59	0.23
1–4	0.58	0.57	0.01	15a–15b	14.30	14.26	0.04
2–3	7.34	7.40	−0.06	15a–16	2.16	2.12	0.04
2–4	1.07	1.12	−0.05	15a–22	0.63	0.47	0.16
3–4	8.06	8.10	−0.04	15b–16	3.87	4.03	−0.16
8–13	10.48	10.10	0.38	20a–20b	14.77	14.76	0.01
11a–11b	17.39	17.26	0.13	20a–22	1.47	1.41	0.06
11a–12	3.27	3.50	−0.23	20a–23a	1.76	1.79	−0.03
11b–12	8.42	8.74	−0.32	20b–27	0.63	0.57	0.06
12–13	3.42	3.48	−0.06	22–23a	6.12	5.64	0.48
13–14	3.07	3.43	−0.36	22–23b	6.74	6.97	−0.23
13–15b	0.54	0.48	0.06	23a–23b	13.57	13.37	0.20
14–15a	1.88	1.94	−0.06				

[a]Calculated for the 96.9–3.1% conformer ratio.

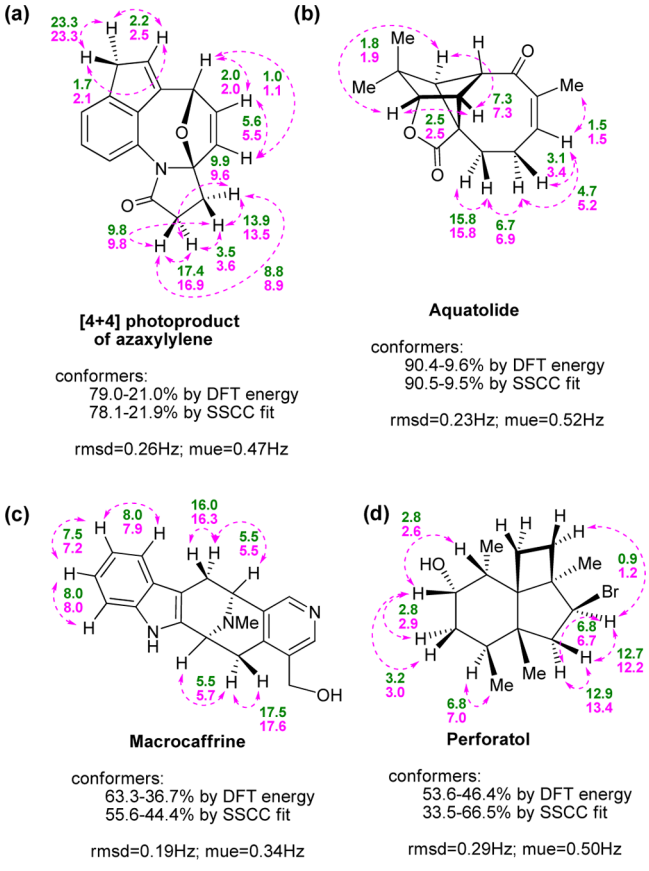


Figure 10. Examples of conformational averaging.

see the Supporting Information). The content of the minor conformer obtained by SSCCs fitting, 21.9%, is in excellent agreement with the value based on the DFT energy, 21%.

Sesquiterpenoid lactone aquatolide, Figure 10b, is a recent example of a structure revision based on theoretical predictions.[23] Our fitting of the calculated constants for its two major conformers predicted 9.5% content of the minor conformer, which is within 0.1% of the value calculated based on the DFT energy difference for two conformers. The predicted SSCCs are well matched with experimental constants, rmsd = 0.23, mue = 0.55 Hz. Similar conformer content, 91.4 and 8.6%, was reported in the original paper,[23] although some constants were predicted with less precision.

The proton coupling constants for alkaloid macrocaffrine[24] were predicted with rmsd = 0.19 Hz and mue = 0.34 Hz, Figure 10c, by fitting two major conformers, 55.6–44.4%. There was almost 10% difference in the conformer content predicted by DFT energy. One notes that the increase of the minor conformer from 36.7% to 44.4%, necessitated by the SSCCs fitting, corresponds to a very small adjustment in the free energy for this equilibrium, $\Delta\Delta G$ < 0.2 kcal/mol, which is beyond DFT accuracy. This example illustrates one of our conclusions that accurate SSCCs may help better understand the actual position of the equilibrium in complex systems, especially in those cases were conformational change results in large changes of magnitude of spin–spin coupling.

The last example in Figure 10d, halogenated sesquiterpene perforatol,[25] further illustrates this point. The major conformer by DFT (53.6%) is predicted to be minor by SSCCs fitting (33.5%). In terms of free energy difference, this also corresponds to a small value $\Delta\Delta G$ < 0.3 kcal/mol at 20 °C, which is not attainable by the DFT methods.

The energy values reported so far are based on B3LYP/6-311+G(d,p) single-point calculations for molecular structures optimized at the B3LYP/6-31G(d) level of theory. The last two examples suggest that it is not a wise investment of time to refine single-point energies at higher levels of theory. It is much more practical to obtain the conformer content at equilibrium by SSCCs fitting and then compare it to the Boltzmann values derived from cheaper/faster calculations at B3LYP/6-31G(d) to make sure that the differences do not exceed the acceptable error of 1–1.5 kcal/mol.

A new *Lycopodium* alkaloid isopalhinine A, Figure 11, recently isolated and characterized by the Zhao group,[26] has two conformationally flexible trimethylene moieties, but only two major conformers accounted for >97% according to DFT energies. Zhao has graciously shared with us high-quality 600 MHz raw NMR data for this alkaloid, from which we accurately measured 27 SSCCs using the Mnova[27] line fitting. Overall, these 27 experimental spin–spin coupling constants matched calculated values for two equilibrating conformers with rmsd = 0.24 Hz and mue = 0.44 Hz. The minor conformer was predicted at 6.5% by DFT, whereas J-fitting required 10.5%. Again, this difference corresponds to a small change in their

dx.doi.org/10.1021/jo501781b | J. Org. Chem. 2014, 79, 8397–8406

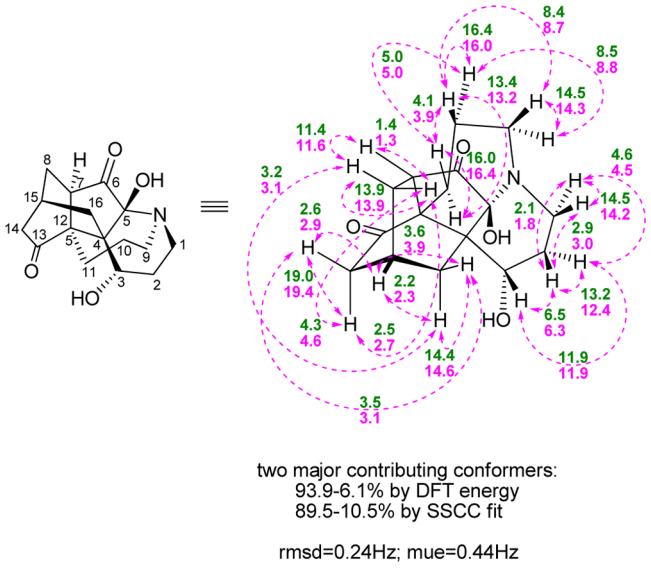


Figure 11. Isopalhinine A.

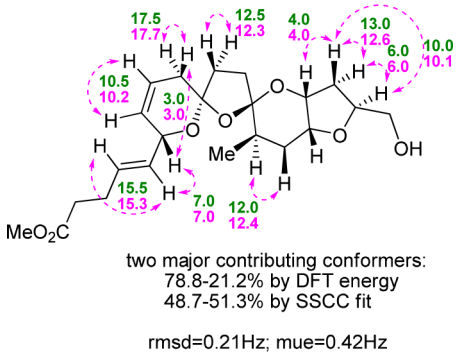


Figure 12. ABCD domain of azaspiracid-1.

relative energy from 1.63 kcal/mol (DFT) to 1.25 kcal/mol (*J* fitting).

Figure 12 illustrates that the method performs well even for a complex conformationally flexible ABCD domain of azaspiracid-1, for which structural revision and total synthesis was performed in the Nicolaou laboratory a decade ago.[28] Since no experimental constants were reported for the conformationally flexible methoxycarbonylethane moiety in the butenyl group attached to ring A, it was simply extended in an all-trans conformation, with no additional conformational averaging. This did not preclude us from obtaining rather small rmsd = 0.21 Hz and mue = 0.42 Hz for the constants which matter for the stereochemical assignment. The same applies to the hydroxymethyl moiety in ring D, for which only one conformation, with the hydroxy group H-bonded to tetrahydrofuranyl oxygen, was used for calculations.

For the 1998 incorrectly assigned stereochemistry of azaspiracid,[29] several predicted constants in rings A and B deviated from the experimental constants considerably, especially for the H−C−O proton in ring A. Even the *cis*-alkene constant, predicted for the wrong structure at 9.8 Hz, deviated significantly more (by 0.7 Hz) from the experimental value of 10.5 Hz, whereas for the correct structure this value was calculated at 10.2 Hz, i.e., within 0.3 Hz of the experimental value.

Finally, we looked at another well-studied alkaloid, morphine. For its free base the nitrogen pyramidal inversion is known, with calculated content of the major conformer (i.e., the methyl group is in equatorial position of the piperidine ring) ranging from 94% by DFT[30] to as low as 71% in an earlier semiempirical study.[31]

We used Neville's experimental NMR data[32] in $CDCl_3$ which we augmented with three small, <2 Hz, constants reported only for the methanol-d_4 solution. Figure 13 shows that the best fit

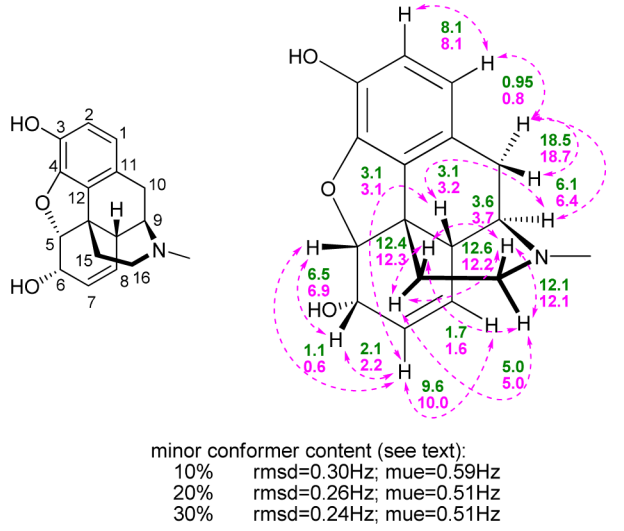


Figure 13. Morphine; best fit SSCCs are shown.

for the SSCCs calculated for the two conformers is achieved when the invertomers are weighted 70−30%. In this case, rmsd = 0.24 Hz and mue = 0.51 Hz. The energy difference required to accommodate the decrease from the 94% content calculated at the B3LYP/6-311++G(d,p) level of DFT theory to 70% is 1.1 kcal/mol, which makes it marginally acceptable but also points to the limits of SSCC fitting because pyramidal inversion at nitrogen in this case exerts only minor effect on SSCCs. The calculated geminal constant at C16 changes from 11.2 Hz (axial-Me) to 14.3 Hz (equatorial-Me), and one of the trans constants at C15−C16 increases by 1 Hz from 11.8 to 12.8 Hz. This small dynamic range makes it difficult to confidently model the exact percentages based solely on the calculated constants for individual conformers because rmsd falls within the acceptable range of 0.2−0.3 Hz for rather broad variations in the invertomer ratios. This is a minor point, as these issues do not preclude an accurate stereochemical assignment.

2.4. Current Challenges. The examples presented so far attest to high fidelity of the method. Of inherent parametrization problems we have identified only one hybridization type with relatively poor linear scaling: $H_2C{=}X$, i.e., the geminal sp^2 moiety, in which the carbon is connected to a heteroatom. This applies mostly to X = nitrogen (as in diazomethane, formaldoxime, and formaldimines), whereas formaldehyde, $H_2C{=}O$ $^2J_{exp}$ = 40.7 Hz, is predicted rather accurately: $^2J_{calc}$ = 40.3 Hz. Partially, this deficiency is due to scarcity and, perhaps, the quality of training experimental constants, which are not easily measured. This hopefully will be resolved in the future as a better training set becomes available. Luckily, it does not constitute a big problem as the relevance of this particular type to accurate structural assignments is at best tangential.

dx.doi.org/10.1021/jo501781b | J. Org. Chem. 2014, 79, 8397−8406

As mentioned above, the largest deviation (1.05 Hz) in the training set involved a vicinal sp^3–sp^3 constant of cyclobutene. It appears that a combination of a CH–CH–C=C– and a cyclobutyl moieties produces somewhat elevated errors. Figure 14 further illustrates this point with dichomitol, a sesquiterpene

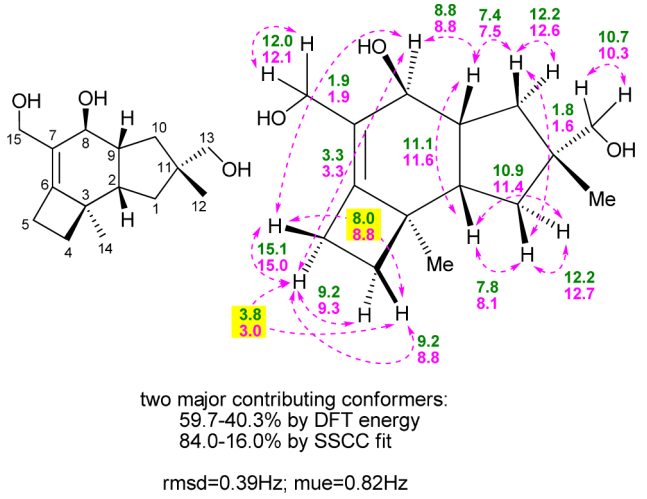


Figure 14. Dichomitol: increased SSCCs errors in the cyclobutane moiety with an sp^2 carbon.

for which the structure was recently corrected by Wei.[33] It is evident that the two vicinal SSCCs for the homoallylic cyclobutane proton C4–H_β are predicted with +0.8 and −0.8 Hz errors (highlighted in yellow). Overall rmsd was 0.39 Hz, on the basis of two major conformers. Removal of the two offending constants does not change the ratio of the conformers but improves rmsd to 0.3 Hz with mue = 0.49 Hz. These two 0.8 Hz errors are likely derived from the limitations of the expedited B3LYP/6-31G(d) structure optimization method. Encouragingly, 0.8 Hz was the largest deviation that we ever encountered in our test cases.

A known issue complicating the accurate calculations of spin–spin coupling constants is the vibrational correction. Rigorous theoretical treatment of it exists, but it is computationally expensive and limited to very small molecules, for example, ammonia.[34] While this topic is beyond the scope of this paper, complications from vibrational mixing are minor in organic molecules. For example, Sneskov[35] calculated vibrational corrections to H–H coupling constants at a very high CCSD/dzp level of theory and reported that they were mostly within 4–8%. Instructively, these contributions varied depending on the hybridization and type of the constants (geminal/vicinal). It is very likely that this problem is mostly mitigated within our parametric model because our scaling is developed with a training set including a large number of points per individual hybridization types and should to a large extent account for the vibrational corrections.

2.5. Misassigned Structures. From the examples presented so far, it is evident that our method offers fast and accurate calculations of proton spin–spin coupling constants, which can greatly facilitate structure assignments, as there is decidedly more structural information in SSCCs than in chemical shifts. For example, the problems with the misassigned structure of paesslerin A,[36] could have been discovered earlier based on just one coupling constant for the alkenic proton (C11–H), which was listed as a broadened singlet but predicted to be a doublet with a 7.5 Hz splitting on the bridgehead proton C6-H. In the synthetic sample synthesized by Ihara[36b] this experimental constant is 7.1 Hz. One concludes that C6 in the actual structure of paesslerin A must be substituted; i.e., it is carrying a methyl or acetoxymethyl group. However, without the raw NMR data it is not possible to make specific predictions.

Another prominent misassigned structure is aldingenin B[37a] for which the total synthesis was accomplished by Crimmins,[37b] who concluded that the originally proposed structure was incorrect. Our predicted spectrum nicely matched Crimmins' synthetic *aldingenin B*, Figure 15, rmsd = 0.22 Hz.

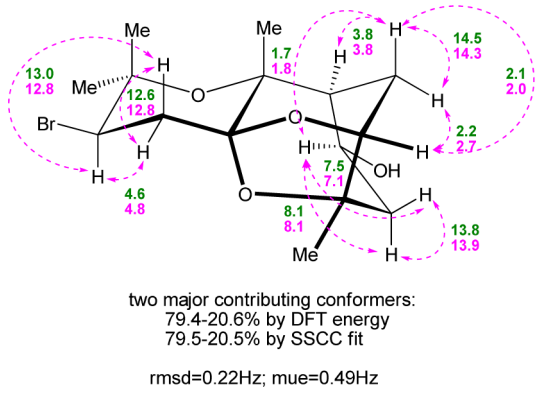


Figure 15. Synthetic aldingenin B.

In the original (mis)assignment of natural aldingenin B,[37a] Lago and coauthors partially draw on their earlier work on a related bisabolene derivative from red algae *Laurencia aldingensis*, aldingenin A.[38] While we are not aware of any attempt to independently synthesize this brominated sesquiterpene, on the basis of our calculations of SSCCs *we are very certain that the structure of aldingenin A is also misassigned*, Figure 16.

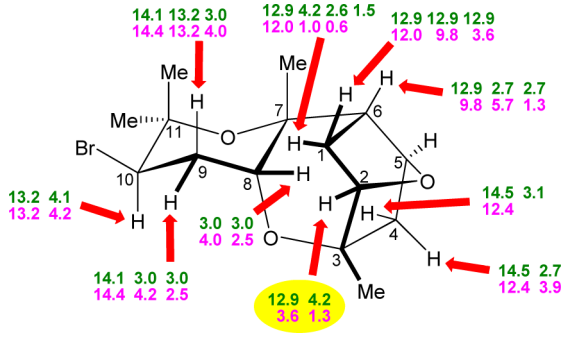


Figure 16. Proposed structure of aldingenin A annotated with experimental (green) and calculated (magenta) SSCCs, listed in descending order.

Judging by the discrepancies between the experimental and calculated spin–spin coupling constants, it appears that the isolated natural product possesses the bromopyran ring as assigned. However, unlike oxachamigrene (Figure 7), aldingenin A does not contain the proposed 7-oxabicyclo[2.2.1]-heptane moiety. The mismatch is particularly evident for proton C2–H, where the largest vicinal constant H–C1–C2–H is reported at 12.9 Hz but calculated at 3.6 Hz (highlighted in yellow). Again, without accurate fitting of the raw NMR data

dx.doi.org/10.1021/jo501781b | J. Org. Chem. 2014, 79, 8397–8406

it is difficult to propose the correct structure. Meanwhile the misassigned structure is propagated in natural product reviews.[39]

At this point, we can only affirm our support of Nicolaou's call for an NMR depository, similar to that for X-ray data at CCDC. Small PDF images of spectra in the Supporting Information serve their role for compliance and quality control but are useless for any practical data mining or refining. Not only will this help with misassigned structures but also will alleviate another commonplace problem, the alarming abundance of typos and typographic errors in published NMR data, even when the structures are correctly assigned.

3. CONCLUSIONS

We have developed a fast, highly accurate method for calculating proton spin−spin coupling constants via a hybridization-discriminated multiparameter scaling scheme for Fermi contacts readily computed with (i) a new DU4 basis set for hydrogen atoms, (ii) an inexpensive 4-31G basis set for carbons, and (iii) B3LYP/6-31G(d) molecular structures. Our approach consistently outperforms the existing methods both in accuracy and computational time, as we illustrated with a number of examples of complex organic synthetic and natural products. With a modest Linux cluster, NMR spectra can be predicted within 1 h for the majority of organic molecules with accuracy fully adequate for an unambiguous stereochemical assignment. Our methodology is systematically improvable as more training data becomes available. There is every reason to believe that such improvements will be similar to the developments in molecular mechanics where more precise and sophisticated force fields have emerged and evolved over time.

■ ASSOCIATED CONTENT

Supporting Information

Computational details, structures, and energies. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: akutatel@du.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Support of this work by the National Science Foundation (CHE-1057800, CHE-1362959) is gratefully acknowledged. We thank Michael Crimmins (UNC), Qin-Shi Zhao (Chinese Academy of Sciences), Martin Banwell (Australian National University), and Xiao-Yi Wei (Chinese Academy of Sciences) for sharing NMR data with us.

■ REFERENCES

(1) Admittedly, even with the limited information contained in chemical shifts, chemists sometimes can work miracles as exemplified by the structure correction of hexacyclinol: Rychnovsky, S. *Org. Lett.* **2006**, *8*, 2895 and references cited therein.

(2) For recent reviews on the theory of spin−spin coupling constants in NMR spectra and utilization of computations for predicting spectra of complex organic compounds, see: (a) Helgaker, T.; Coriani, S.; Jorgensen, P.; Kristensen, K.; Olsen, J.; Ruud, K. *Chem. Rev.* **2012**, *112*, 543. (b) Rusakov, Yu. Yu.; Krivdin, L. B. *Russ. Chem. Rev.* **2013**, *82*, 99. (c) Bifulco, G.; Dambruoso, P.; Gomez-Paloma, L.; Riccio, R. *Chem. Rev.* **2007**, *107*, 3744.

(3) (a) Onak, T.; Jaballas, J.; Barfield, M. *J. Am. Chem. Soc.* **1999**, *121*, 2850. (b) Scheurer, C.; Bruschweiler, R. *J. Am. Chem. Soc.* **1999**, *121*, 8661. (c) Del Bene, J. E.; Bartlett, R. J. *J. Am. Chem. Soc.* **2000**, *122*, 10480. (d) Del Bene, J. E.; Perera, S. A.; Bartlett, R. J. *J. Am. Chem. Soc.* **2000**, *122*, 3560. (e) Del Bene, J. E.; Perera, S. A.; Bartlett, R. J. *J. Phys. Chem. A* **2001**, *105*, 930. (f) Diez, E.; Casanueva, J.; San Fabian, J.; Esteban, A. L.; Galache, M. P.; Barone, V.; Peralta, J. E.; Contreras, R. H. *Mol. Phys.* **2005**, *103*, 1307.

(4) Jensen, F. *J. Chem. Theory Comput.* **2006**, *2*, 1360.

(5) Bally, T.; Rablen, P. R. *J. Org. Chem.* **2011**, *76*, 4818.

(6) Lopez-Vallejo, F.; Fragoso-Serrano, M.; Suarez-Ortiz, G. A.; Hernandez-Rojas, A. C.; Cerda-García-Rojas, C. M.; Pereda-Miranda, R. *J. Org. Chem.* **2011**, *76*, 6057.

(7) Wilkens, S. J.; Westler, W. M.; Markley, J. L.; Weinhold, F. *J. Am. Chem. Soc.* **2001**, *123*, 12026.

(8) (a) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211. (b) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1983**, *78*, 4066. (c) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735.

(9) Peralta, J. E.; Contreras, R. H.; Snyder, J. P. *Chem. Commun.* **2000**, 2025.

(10) The latest version of NBO program: NBO 6.0: Glendening, E. D.; Badenhoop, K.; Reed, A. E. ; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Landis, C. R.; Weinhold, F. Theoretical Chemistry Institute, University of Wisconsin, Madison, 2013.

(11) (a) Kutateladze, A. G.; McHale, W. A., Jr. *ARKIVOC* **2005**, *IV*, 88. (b) Kutateladze, A. G. *J. Am. Chem. Soc.* **2001**, *123* (38), 9279.

(12) National Institute for Advanced Industrial Science and Technology, Japan, http://sdbs.db.aist.go.jp/sdbs/cgi-bin/cre_index.cgi.

(13) (a) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *J. Cheminf.* **2011**, *3*, 33. (b) http://openbabel.org.

(14) Valiulin, R. A.; Arisco, T. M.; Kutateladze, A. G. *Org. Lett.* **2010**, *12*, 3398.

(15) Based on a large in-house database of experimental NMR spectra, we compute chemical shifts in chloroform at the mPW1PW91/6-311+G(d,p) level of theory and scale them with a weakly curved empirical cubic equation: $-0.002249*\delta^3 + 0.1714*\delta^2 - 5.239*\delta + 65.55$, which includes the TMS reference value for stability. The obtained values of chemical shifts in ppm can be further linearly corrected with experimental chemical shifts.

(16) (a) Bos, P. H.; Antalek, M. T.; Porco, J. A.; Stephenson, C. R. J. *J. Am. Chem. Soc.* **2013**, *135*, 17978. (b) Compound numbers **22** and **50** are from the paper described in ref 16a.

(17) Corey, E. J.; Kang, M. C.; Desai, M. C.; Ghosh, A. K.; Houpis, I. N. *J. Am. Chem. Soc.* **1988**, *110*, 649.

(18) Brito, I.; Cueto, M.; Diaz-Marrero, A. R.; Darias, J.; San Martin, A. S. *J. Nat. Prod.* **2002**, *65*, 946.

(19) Okamura, H.; Iwagawa, T.; Nakatani, M. *Bull. Chem. Soc. Jpn.* **1995**, *68*, 3465.

(20) Strychnine SSCCs are averaged from a compilation in: Cobas, J. C.; Constantino-Castillo, V.; Martin-Pastor, M.; del Rio-Portilla, F. *Magn. Reson. Chem.* **2012**, *50*, S86.

(21) Butts, C. P.; Jones, C. R.; Harvey, J. N. *Chem. Commun.* **2011**, *47*, 1193.

(22) Mukhina, O. A.; Kumar, N. N. B.; Arisco, T. M.; Valiulin, R. A.; Metzel, G. A.; Kutateladze, A. G. *Angew. Chem., Int. Ed.* **2011**, *50*, 9423.

(23) Lodewyk, M. W.; Soldi, C.; Jones, P. B.; Olmstead, M. M.; Rita, J.; Shaw, J. T.; Tantillo, D. J. *J. Am. Chem. Soc.* **2012**, *134*, 18550.

(24) (a) Ohba, M.; Natsutani, I. *Tetrahedron* **2007**, *63*, 12689. (b) Original discovery: Nasser, A. M. A. G.; Court, W. E. *Phytochemistry* **1983**, *22*, 2297.

(25) Findlay, J. A.; Li, G. *Can. J. Chem.* **2002**, *80*, 1697.

(26) Dong, L.-B.; Gao, X.; Liu, F.; He, J.; Wu, X.-D.; Li, Y.; Zhao, Q.-S. *Org. Lett.* **2013**, *15*, 3570.

(27) As implemented in MestReNova software, version 9.0.0, 2013 Mestrelab Research S.L., http://www.mestrelab.com.

(28) Nicolaou, K. C.; Koftis, T. V.; Vyskocil, S.; Petrovic, G.; Ling, T.; Yamada, Y. M. A.; Tang, W.; Frederick, M. O. *Angew. Chem., Int. Ed.* **2004**, *43*, 4318.
(29) Satake, M.; Ofuji, K.; Naoki, H.; James, K. J.; Furey, A.; McMahon, T.; Silke, J.; Yasumoto, T. *J. Am. Chem. Soc.* **1998**, *120*, 9967.
(30) Based on the reported 6.8 kJ/mol energy difference with the second conformer: Baranska, M.; Kaczor, A. *J. Raman Spectrosc.* **2012**, *43*, 102.
(31) Goldblum, A.; Loew, G. H. *Eur. J. Pharmacol., Mol. Pharm.* **1991**, *206*, 119.
(32) Neville, G. A.; Ekiel, I.; Smith, I. C. P. *Magn. Reson. Chem.* **1987**, *25*, 31.
(33) (a) Original report of a triquinane structure of dichomitol (misassigned): Huang, Z.; Dan, Y.; Huang, Y.; Lin, L.; Li, T.; Ye, W.; Wei, X. *J. Nat. Prod.* **2004**, *67*, 2121. (b) Synthesis of the originally proposed structure: Mehta, G.; Pallavi, K. *Tetrahedron Lett.* **2006**, *47*, 8355. (c) Structure correction: Xie, H.-H.; Xu, X.-Y.; Dan, Y.; Wei, X.-Y. *Helv. Chim. Acta* **2011**, *94*, 868.
(34) Yachmenev, A.; Yurchenko, S. N.; Paidarová, I.; Jensen, P.; Thiel, W.; Sauer, S. P. A. *J. Chem. Phys.* **2010**, *132*, 114305.
(35) Sneskov, K.; Stanton, J. F. *Mol. Physics* **2012**, *110*, 2321.
(36) (a) Original incorrect assignment: Brasco, M. F. R.; Seldes, A. M.; Palermo, J. A. *Org. Lett.* **2001**, *3*, 1415. (b) Synthesis of the proposed structure and conclusion of misassignment: Inanaga, K.; Takasu, K.; Ihara, M. *J. Am. Chem. Soc.* **2004**, *126*, 1352.
(37) (a) Original incorrect assignment: de Carvalho, L. R.; Fujii, M. T.; Roque, N. F.; Lago, J. H. G. *Phytochemistry* **2006**, *67*, 1331. (b) Synthesis of the proposed structure and conclusion of misassignment: Crimmins, M. T.; Hughes, C. O. *Org. Lett.* **2012**, *14*, 2168.
(38) de Carvalho, L. R.; Fujii, M. T.; Roque, N. F.; Kato, M. J.; Lago, J. H. G. *Tetrahedron Lett.* **2003**, *44*, 2637.
(39) (a) Blunk, J. W.; Copp, B. R.; Munro, M. H. G.; Northcote, P. T.; Prinsep, M. R. *Nat. Prod. Rep.* **2005**, *22*, 15. (b) Fraga, B. M. *Nat. Prod. Rep.* **2004**, *21*, 669.